

METHOD AND APPARATUS FOR A MULTI-GIGABIT ETHERNET ARCHITECTURE

Publication number: JP2003500899 (T)

Publication date: 2003-01-07

Inventor(s):

Applicant(s):

Classification:

- **international:** *H04L12/28; H04L12/40; H04L12/413; H04L29/06; H04L29/08; H04L12/28; H04L12/40; H04L12/407; H04L29/06; H04L29/08; (IPC1-7): H04L12/28; H04L12/40*

- **European:** *H04L29/06*

Application number: JP20000619163T 20000517

Priority number(s): US19990314782 19990519; WO2000US13584 20000517

Also published as:

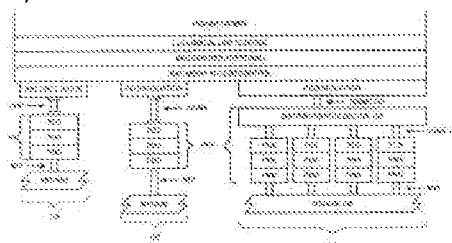
WO0070827 (A2)
WO0070827 (A3)
US6873630 (B1)
EP1179248 (A2)
EP1179248 (B1)

[more >>](#)

Abstract not available for JP 2003500899 (T)

Abstract of corresponding document: **WO 0070827 (A2)**

An Ethernet architecture enables the transfer of data by striping individual frames across a plurality of logical channels, thus allowing operation at substantially the sum of the individual channel rates. A distributor within a sending entity's network interface distributes frame bytes in a round-robin fashion on the plurality of channels. Each mini-frame is separately framed and encoded for transmission across its channel. A receiving entity's network interface includes a collector for collecting multiple mini-frames and reconstructing the frame's byte stream. The first and last bytes of each frame and mini-frame are marked for ease of recognition. Multiple unique idle symbols may be employed for transmission during inter-frame gaps to facilitate the collector's synchronization of the multiple channels and/or enhance error detection.; A maximum channel skew is specified, and each channel may be buffered with an elasticity proportional to the maximum skew so that propagation delay may be encountered between channels without disrupting communications.



.....
Data supplied from the **espacenet** database — Worldwide

(19)日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11)特許出願公表番号
特表2003-500899
(P2003-500899A)

(43)公表日 平成15年1月7日(2003.1.7)

(51)Int.Cl. ⁷	識別記号	F I	テームコード* (参考)
H 0 4 L 12/28	2 0 0	H 0 4 L 12/28	2 0 0 Z 5 K 0 3 2
12/40		12/40	Z 5 K 0 3 3

審査請求 未請求 予備審査請求 有 (全 50 頁)

(21)出願番号 特願2000-619163(P2000-619163)
(86) (22)出願日 平成12年5月17日(2000.5.17)
(85)翻訳文提出日 平成13年11月16日(2001.11.16)
(86)国際出願番号 PCT/US 00/13584
(87)国際公開番号 WO 00/070827
(87)国際公開日 平成12年11月23日(2000.11.23)
(31)優先権主張番号 09/314,782
(32)優先日 平成11年5月19日(1999.5.19)
(33)優先権主張国 米国 (US)

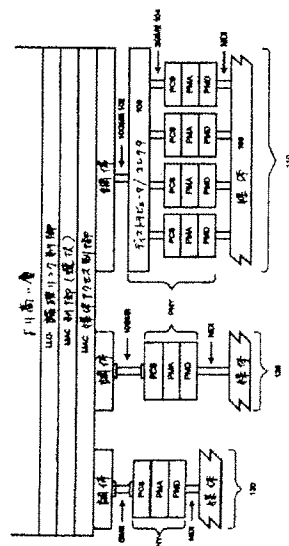
(71)出願人 サン・マイクロシステムズ・インコーポレ
イテッド
Sun Microsystems, I
nc.
アメリカ合衆国 カリフォルニア 95054,
サンタ クララ, ネットワーク サー
クル 4150
(72)発明者 ムラー, シモン
アメリカ合衆国 カリフォルニア 94086,
サニーバール, ラ メサ テラス
983, アパートメント ディー
(74)代理人 弁理士 山本 秀策

最終頁に続く

(54)【発明の名称】 複数ギガビットイーサネット (R) アーキテクチャの方法および装置

(57)【要約】

イーサネット (R) アーキテクチャは、複数の論理チャネルを介して個々のフレームをストライピングすることによってデータ転送を可能にする。それ故、実質的に個々のチャネルの合計速度での動作が可能になる。送信側エンティティのネットワークインターフェースにおけるディストリビュータは、フレームバイトをラウンドロビン方式で複数のチャネル上に分配する。それぞれのミニフレームは、それぞれのチャネルにわたる伝送のために別個にフレーム化およびエンコードされる。受信側エンティティのネットワークのインターフェースは、複数のミニフレームを収集し、フレームのバイトストリームを再構築するコレクタを含む。



【特許請求の範囲】

【請求項1】 複数のチャネルを介して、第1ネットワークエンティティから第2ネットワークエンティティに通信を伝送する方法であって、

第1ネットワークエンティティにて第2ネットワークエンティティに対する通信を受信する工程と、

該第1ネットワークエンティティを該第2ネットワークエンティティに連結する複数のチャネルのそれぞれを介して、同期化情報をブロードキャストする工程と、

該通信を複数の部分に分割する工程と、

ミニフレームとして該複数の部分のそれぞれをエンコードする工程であって、該ミニフレームのそれぞれが開始要素および終了要素を含む、エンコードする工程と、

第1ミニフレームを該複数のチャネルの第1チャネルで送信する工程と、

第2ミニフレームを該複数のチャネルの第2チャネルで送信する工程と、を包含する、方法。

【請求項2】 前記通信は、前記複数のチャネルを介して、前記第2エンティティに1秒あたり1ギガビットより速いデータ速度で伝送される、請求項1に記載の方法。

【請求項3】 前記通信がイーサネット（R）フレームであり、該通信の複数の部分のそれぞれが1バイト以上を含む、請求項1に記載の方法。

【請求項4】 前記エンコードする工程が、

前記通信の第1部分の第1要素をパケットデリミッターの開始としてエンコードする工程と、

該通信の第2部分の第1要素をミニフレームデリミッターの開始としてエンコードする工程と、

を包含する、請求項3に記載の方法。

【請求項5】 前記通信の複数の部分のそれぞれが最小の数の前記プリアンブルバイトを含むことを確実にするために、前記分割する工程の前に前記イーサネット（R）フレームのプリアンブルバイトの数を増やす工程をさらに包含する

、請求項3に記載の方法。

【請求項6】 前記受信する工程が、1秒あたり1ギガビットより速いデータ速度で前記通信を伝達するように構成された第1インターフェースを介して、媒体アクセス制御モジュールから、ネットワークインターフェースデバイスの分配モジュールにて、通信を受信する工程を包含する、請求項1に記載の方法。

【請求項7】 前記第1ミニフレームを送信する工程は、前記通信の第1部分を第1物理的コーディングモジュールに第2インターフェースを介して転送する工程を包含し、

該第1物理的コーディングモジュールが該通信の第1部分を前記第1チャンネル上で伝送するための一連のコードにエンコードするように構成される、請求項6に記載の方法。

【請求項8】 前記エンコードする工程が、

前記第1部分の第1要素が前記通信の第1要素である場合には、第1開始コードで該第1部分の第1要素をエンコードし、そうでない場合に、第2開始コードで該第1部分の第1要素をエンコードする工程と、

該第1部分の最後の要素が該通信の最後の要素である場合には、第1終了コードで該第1部分の最後の要素をエンコードし、そうでない場合には、第2終了コードで該第1部分の最後の要素をエンコードする工程と、
を包含する、請求項7に記載の方法。

【請求項9】 前記第2インターフェースが、前記第1部分を1秒あたり1ギガビットより速いデータ速度で伝達するように構成される、請求項7に記載の方法。

【請求項10】 前記分割する工程が前記複数のチャンネル間で前記通信の要素を割り当てる工程を包含する、請求項1に記載の方法。

【請求項11】 前記複数のチャンネルのそれぞれが別個の物理的通信リンクにまたがるように構成される、請求項10に記載の方法。

【請求項12】 前記複数のチャンネルのそれぞれが共通の物理的通信リンクにまたがるように構成される、請求項10に記載の方法。

【請求項13】 前記第1ミニフレームおよび前記第2ミニフレームの開始

要素の一方が、前記通信の開始を示すように構成された第1開始記号であり、該第1ミニフレームおよび該第2ミニフレームの他の開始要素が、該通信の一部の開始を示すように構成された第2開始記号であり、

該第1ミニフレームおよび該第2ミニフレームの終了要素の一方が、該通信の終了を示すように構成された第1終了記号であり、該第1ミニフレームおよび該第2ミニフレームの他の終了要素が、該通信の一部の終了を示すように構成された第2終了記号である、請求項1に記載の方法。

【請求項14】 前記ブロードキャストする工程が、前記第1チャンネルおよび前記第2チャンネル上で第1アイドル信号を伝送する工程を包含し、

前記方法が、最終ミニフレームの送信後に該第1チャンネルおよび該第2チャンネル上で該第1アイドル信号とは異なる第2アイドル信号を伝送する工程をさらに包含する、請求項1に記載の方法。

【請求項15】 前記エンコードする工程が、

第1開始デリミッターで前記第1ミニフレームに相当する前記通信の第1部分の第1要素をエンコードする工程と、

第2開始デリミッターで前記第2ミニフレームに相当する該通信の第2部分の第1要素をエンコードする工程と、
を包含する、請求項1に記載の方法。

【請求項16】 前記エンコードする工程が、

第1終了デリミッターで前記通信の前記第1部分の最後の要素をエンコードする工程と、

第2終了デリミッターで該通信の前記第2部分の最後の要素をエンコードする工程と、
をさらに包含する、請求項15に記載の方法。

【請求項17】 第2ネットワークエンティティにおいて第1ネットワークエンティティから、複数のチャンネルを介して、通信を受信する方法であって、

該第1ネットワークエンティティを該第2ネットワークエンティティに連結する複数のチャンネルのそれぞれを介して、第2ネットワークエンティティにて同期化情報を受信する工程と、

該複数のチャネルのそれぞれで、該第1ネットワークエンティティから該第2ネットワークエンティティへの通信のフレーム化された部分を受信する工程と、
該フレーム化された部分のそれぞれで開始要素および終了要素を検出する工程と、
該通信の該フレーム化された部分のそれぞれをデコードする工程と、
該デコードされ、フレーム化された部分から該通信を再セットみ立てする工程と、
を包含する、方法。

【請求項18】 前記通信がイーサネット（R）フレームである、請求項17に記載の方法。

【請求項19】 前記デコードされ、フレーム化された部分のうちの任意の部分におけるバイトの数が、該デコードされ、フレーム化された部分の任意の他の部分におけるバイトの数と1より多く異なる場合に、エラーを示す工程をさらに包含する、請求項18に記載の方法。

【請求項20】 前記フレーム化された部分を受信する工程が、第1チャネルを介して前記第1ネットワークエンティティから第1伝送を受信する工程を包含し、

該第1伝送が、

通信の開始および通信の第1部分の開始とのうちの一方を示すように構成された第1信号と、

該通信の第1部分とを含む、請求項18に記載の方法。

【請求項21】 前記フレーム化された部分を受信する工程が、第2チャネルを介して前記第1ネットワークエンティティから第2伝送を受信する工程をさらに包含し、

該第2伝送が、

該通信の第2部分と、

該通信の終了の該通信の第2部分の終了との一方を示すように構成された第2信号とを含む、請求項20に記載の方法。

【請求項22】 前記複数のチャネルのそれぞれが共通の通信媒体にまたが

る、請求項17に記載の方法。

【請求項23】 前記複数のチャンネルのそれぞれが別個の物理的媒体にまたがる、請求項17に記載の方法。

【請求項24】 前記再セットみ立てする工程が、前記通信の第1のデコードされ、フレーム化された部分の要素を該通信の第2のデコードされ、フレーム化された部分の要素と結合させる工程を包含する、請求項17に記載の方法。

【請求項25】 前記結合された要素を媒体アクセス制御モジュールに、該結合された要素を伝達するように構成された第1インターフェースを介して、1秒あたり1ギガビットより速いデータ速度で送信する工程をさらに包含する、請求項24に記載の方法。

【請求項26】 同期化情報を受信する工程が、前記通信のフレーム化された部分を受信する工程の前に前記複数のチャンネルのそれぞれで第1アイドルコードを受信する工程を包含し、

該方法が、前記再セットみ立てする工程の後に、該複数のチャンネルのそれぞれで第2アイドルコードを受信する工程をさらに包含する、請求項17に記載の方法。

【請求項27】 第1パケットを複数の通信チャンネル上で分配するように構成されたディストリビュータと、

該第1パケットの第1サブセットを第1通信チャンネル上で伝送するためにエンコードするように構成された第1物理的コーディングモジュールであって、該第1のエンコードされたサブセットが第1開始要素および第1終了要素を含む、第1物理的コーディングモジュールと、

該第1パケットの第2サブセットを第2通信チャンネル上で伝送するためにエンコードするように構成された第2物理的コーディングモジュールであって、該第2のエンコードされたサブセットが第2開始要素および第2終了要素を含む、第2物理的コーディングモジュールと、
を含む、コンピュータシステムとネットワークとの間にインターフェースをとるネットワークインターフェースデバイス。

【請求項28】 第2パケットを前記複数の通信チャンネルを介して受信する

ように構成されたコレクタをさらに含むネットワークインターフェースデバイスであって、

前記第1物理的コーディングモジュールが、第3開始要素および第3終了要素を含む、前記第1通信チャネルを介して受信された該第2パケットの第1のエンコードされたサブセットをデコードするようにさらに構成され、

前記第2物理的コーディングモジュールが、第4開始要素および第4終了要素を含む、前記第2通信チャネルを介して受信された該第2パケットの第2のエンコードされたサブセットをデコードするようにさらに構成される、請求項27に記載のネットワークインターフェースデバイス。

【請求項29】 クロック信号の両端で同期化する際に、前記ディストリビュータおよび前記コレクタと媒体アクセス制御モジュールとの間に1秒あたり1ギガビットより速いデータ速度でインターフェースをとるように構成された、第1インターフェースと、

第2インターフェースのセットとであって、該第2インターフェースのそれぞれが、第2クロック信号の両端と同期化する際に、該ディストリビュータおよび該コレクタと前記物理的コーディングモジュールの1つとの間に1秒あたり1ギガビットより速いデータ速度でインターフェースをとるように構成された、第2インターフェースのセットと、

をさらに備える、請求項27に記載のネットワークインターフェースデバイス。

【請求項30】 前記第1インターフェースが、前記第2インターフェースのセットの動作速度の合計とほぼ同等のデータ速度で動作するように構成された、請求項29に記載のネットワークインターフェースデバイス。

【請求項31】 前記第1インターフェースが、1秒あたり約10ギガビットのデータ速度で動作するように構成された、請求項29に記載のネットワークインターフェースデバイス。

【請求項32】 前記通信チャネルのそれぞれを介して受信された前記第2パケットのバイト数を比較するように構成された、エラー検出器をさらに含み、

該通信チャネルの1つで受信された該第2パケットのバイトの数が、該通信チャネルのうちの第2のチャネルを介して受信された該第2パケットのバイト数と

1 バイトより多く異なる場合に、該エラー検出器がエラーを示すようにさらに構成される、請求項28に記載のネットワークインターフェースデバイス。

【請求項33】 コンピュータによって実行される場合、第1ネットワークエンティティから第2ネットワークエンティティに複数のチャネルを介して通信を分配する方法をコンピュータに実施させる命令を格納するコンピュータ読み出し可能な記憶媒体であって、該方法が、

該第1ネットワークエンティティにて該第2ネットワークエンティティのための通信を受信する工程と、

該第1ネットワークエンティティを該第2ネットワークエンティティに連結させる複数のチャネルのそれぞれを介して同期化情報をブロードキャストする工程と、

該通信を複数の部分に分割する工程と、

該複数の部分のそれぞれをミニフレームとしてエンコードする工程であって、該ミニフレームのそれぞれが開始要素および終了要素を含む、工程と、

第1ミニフレームを該複数のチャネルの第1チャネルに送信する工程と、

第2ミニフレームを該複数のチャネルの第2チャネルに送信する工程と、を包含する、コンピュータ読み出し可能な記憶媒体。

【請求項34】 コンピュータによって実行される場合、第2ネットワークエンティティにおいて第1ネットワークエンティティから複数のチャネルを介して通信を受信する方法をコンピュータに実施させる命令を格納するコンピュータで読み出し可能な記憶媒体であって、該方法が、

該第2ネットワークエンティティにて、該第1ネットワークエンティティを該第2ネットワークエンティティに連結させる複数のチャネルのそれぞれを介して同期化情報を受信する工程と、

該複数のチャネルのそれぞれで、該第1ネットワークエンティティから該第2ネットワークエンティティへの通信のフレーム化された部分を受信する工程と、

前記フレーム化された部分のそれぞれで開始要素および終了要素を検出する工程と、

該通信の該フレーム化された部分のそれぞれをデコードする工程と、

該デコードされ、フレーム化された部分から該通信を再セットみ立てする工程と、

を包含する、コンピュータ読み出し可能な記憶媒体。

【発明の詳細な説明】**【0001】****(背景)**

本発明は、コンピュータシステムおよびネットワークの分野に関する。より詳細には、高速のデータ転送でコンピュータシステムまたはその他のデバイスとイーサネット（R）ネットワークとの間にインターフェースをとる方法と装置が提示される。

【0002】

コンピュータシステムは主に、愛好家および専門家の興味の対象であったものから、人口の大半にとって不可欠なツールに発展してきた。コンピュータシステムの数および能力双方の増大に伴い、コンピュータシステム間で通信を行う必要性も増え続けている。周辺機器を共有し、電子メールを配信する初期の使用から、今日流通したアプリケーションおよびクライアント／サーバアーキテクチャの使用まで、コンピュータ通信を配信するネットワークは、サイズおよび範囲において急速に発展してきている。

【0003】

1つの特別なネットワークアーキテクチャであるイーサネット（R）は、ネットワーク伝送速度が急激に増加しても、多くのコンピュータ環境で優位のままである。かつては10Mbpsの通信速度が高速イーサネット（R）ローカルエリアネットワーク（LAN）の標準であったが、今日では100倍高速な（すなわち、1Gbps）イーサネット（R）ネットワークを入手およびインストールすることが可能である。特に、IEEE（Institute of Electrical and Electronics Engineers）802.3規格は、そのようなネットワークの許容データリンクプロトコルを詳細に規定している。

【0004】

ちょうど今日の高速なネットワークに対する明らかに必要であったように、さらに高速な伝送速度を有するネットワークが利用可能になれば直ちに実施される

ことは疑いない。現在1 G b p s（およびより低速な）ネットワーク上で通信するコンピュータシステムおよびアプリケーションと同様、新たなコンピュータシステムおよびアプリケーションが、G b p sの何倍もの速度で動作するネットワークを有意義に利用することは確実であり得る。より高いバンド幅を好む可能性が高い所定のタイプの動作は、大容量のデータを必要とするか、または生成する、マルチメディア、データベース、モデリング、およびその他の分野を含む。

【0005】

例えば、コンピュータシステム「クラスタ」およびその他のよく相互接続されるコンピュータシステムは、より高速な通信速度から大きな利益を受ける。特に、そのようなクラスタにおける計算および動作はしばしば、複数のエンドノード間で共有または分配されるため、迅速なネットワーク通信を所望する度合いは、ノードの内部動作速度（例えば、クラスタメンバーのC P Uが内蔵メモリと通信する速度）によってのみ制限され得る。これらのタイプのネットワークを移行する通信はしばしば、（例えば、相対的に優先度低のユーザレベルではなく）優先度高のシステムレベルで実施されるため、通信が高速で伝達されればされるほど、システムがユーザアクティビティに充てなければならない時間が多くなる。

【0006】

MAN (Metropolitan Area Network)、WAN (Wide Area Network) またはRAN (Regional Area Network) など、クラスタまたはLAN以外のネットワークで動作するアプリケーションもまた、増加した伝送速度から利益を受け得る。しかしながら、これらのタイプのネットワークでは、アプリケーションは、例えばコンピュータクラスタ内のアプリケーションより、ずっと長い距離を介して通信を行う。

【0007】

それ故、1 G b p sより速い伝送速度で動作可能なネットワークアーキテクチャが必要である。特に、インターフェースが1 G b p sより速い速度でネットワークトラフィックを伝えることができるような、コンピュータシステムまたはその他のネットワークエンティティをネットワークにインタフェースする手段が必

要である。イーサネット（R）プロトコルを利用するネットワークおよびネットワークコンポーネントの多さ、ならびにプログラマ、開発者および設計者のこの技術の普及度のために、イーサネット（R）を用いるネットワークを実施することは非常に有益である。1 G b p sより速く動作するイーサネット（R）ネットワークインターフェースは、好適には、すべてではないにしても、ほとんどの既存のイーサネット（R）インプリメンテーションと互換性を有する。インターフェースは好適には、短い距離を介して動作し得るコンピュータクラスタ、およびより長く、地域にさえもわたる距離で動作するネットワークなどの環境に適する。

【0008】

（要旨）

本発明の1実施形態において、コンピュータシステムまたはその他のネットワークエンティティとイーサネット（R）ネットワークとの間にインターフェースをとり、そして1秒あたり複数ギガビットでデータをエンティティ間で伝送する、システムおよび方法が提供される。

【0009】

本実施形態におけるイーサネット（R）ネットワークは、事実上あらゆるタイプの媒体（例えば、ファイバ、ワイヤ）からなる1つ以上の物理的リンクを含む。しかしながら、例えば、おそらく低速通信で動作する場合を除いては、通信が動作の全二重モードでのみ実施されるように、ネットワークが専用モードで動作する。

【0010】

ネットワーク上で通信を取り交わすネットワークエンティティはそれぞれ、通信をネットワークへ挿入しネットワークから取り外すネットワークインターフェースをセットみ込む。本発明の1実施形態におけるネットワークインターフェースは、1つ以上の集積回路、プリント基板、ソフトウェアモジュールなどを含み得る。

【0011】

通信が第1ネットワークエンティティによって、ネットワークを介して伝送さ

れるとき、そのインターフェースは通信を複数の論理チャンネルに分割する。それぞれのチャンネルは、別個の光ファイバまたは有線ケーブルなどの異なる物理的リンクを、あるいは動作の波形分割多重伝送方式（WDM）モードを採用したファイバなどの共通物理的リンクを、移行（t r a n s i t）し得る。受信側エンティティにおけるネットワークインターフェースは、複数チャンネルを受信して、エンティティに転送するためにこれらの複数チャンネルを再セットみ立てする。

【0012】

本発明の1実施形態において、通信は動作の媒体アクセス制御（MAC）層より低いポイントにおいて、複数チャンネルにわたる伝送のために分割される。それ故、本実施形態において、通信のそれぞれのフレームまたはパケットの個々のバイトは、分離され、ラウンドロビン方式でチャンネルのうちの1つを介して送信される。イーサネット（R）ネットワークを介した通信の伝送速度は、それ故、それぞれのチャンネルの速度合計に近似する。本発明のある特定の実施形態において、4つの論理チャンネルが採用される。ここでそれぞれのチャンネルは、通信に10 G b p s の伝送速度を維持するために、約2.5 G b p s で動作する。

【0013】

例えば、イーサネット（R）フレームのそれぞれのミニフレーム（すなわち、1チャンネルによって運搬されるフレーム部分）は、他のミニフレームのサイズ、プラスマイナス1バイトに等しい。これは、フレームの伝送または受信におけるエラーを検出する容易な方法を提供する。さらに、フレームのシーケンシングは、フレーム間の期間（例えば、パケット間ギャップまたはI P G）を表す複数の異なるコードまたは記号を提供することによって実施され得る。このフレームのシーケンシングの方法を用いて、受信側エンティティは、いずれのコードまたは記号がそれぞれのギャップの間に受信されるかをモニタリングすることによって、複数チャンネルの同期を取り得る。

【0014】

本発明の1実施形態において、ネットワークインターフェースの受信において、それぞれのチャンネルに対するバッファが維持される。バッファのサイズは予測されるチャンネルのスキュー（例えば、チャンネルをわたる伝搬時間の差）の最大量

に比例し得る。

【0015】

(詳細な説明)

以下の記載は、すべての当業者が本発明を作成および使用することを可能とするために提示され、本発明の特定の用途および必要条件のコンテキストにおいて提供される。開示された実施形態の種々の変形例が、当業者に理解され、本明細書で規定される概括的な原理は、本発明の精神および範囲から逸脱することなく、他の実施形態および用途に適用され得る。それ故、本発明は例示の実施形態に限定されるものではなく、本明細書に開示される原理および特徴と一致する最も広い範囲が与えられる。

【0016】

特に、高速イーサネット（R）ネットワークインターフェースを実施する装置および関連の方法が提供される。そのようなインターフェースは、例えば、イーサネット（R）ネットワークに連結されたコンピュータシステムまたは他の通信デバイスに適する。通信デバイスが連結されるイーサネット（R）ネットワークの構築に関して、本発明が制限されないことを当業者は理解する。あるネットワークエンティティから別のネットワークエンティティに信号を配信する他の手段が適しているのと同様、1つ以上の光ファイバまたは導電体から構築されたネットワークも、適している。

【0017】

本発明の本実施形態が実施されるプログラム環境は、例えば、汎用コンピュータ、または手持ち式のコンピュータなど専用デバイスをセットみ込む。そのようなデバイス（例えば、プロセッサ、メモリ、データ記憶装置およびディスプレイ）の詳細は、周知であるから、明瞭にするためここでは省略される。

【0018】

また、本発明の技術が、種々の技術を使用して実施され得ることが理解されるべきである。特に、本明細書において記載される方法は、コンピュータシステム上で動作するソフトウェアにおいて実施されてもよいし、マイクロプロセッサのセット合せ、またはその他の特別に設計された特定用途集積回路、プログラマブ

ル論理デバイス、またはこれらの種々のセット合せのいずれかを使用したハードウェアにおいて実施してもよい。本発明の形態または範囲をまったく制限しない、単なる1実施例として、本明細書に記載される方法は、搬送波、ディスクドライブまたはコンピュータで読み出し可能な媒体などの記憶媒体に存在する一連のコンピュータで実行可能な命令に関連して実施され得る。搬送波の例示的形態は、デジタルデータストリームを、ローカルネットワークまたはインターネットなどの一般にアクセス可能なネットワークに沿って伝達する、電気信号、電磁信号または光信号の形態を取り得る。

【0019】

本発明の1実施形態において、1 G b p s より速いデータ転送速度で、コンピュータシステムをイーサネット（R）ネットワークに接続するインターフェースが記載される。本実施形態のある具体的な実施例において、ネットワークインターフェースは、約10 G b p s の速度でイーサネット（R）ネットワークと通信を取り交わす。

【0020】

本実施形態において、イーサネット（R）ネットワークはコンピュータシステムと専用の構成を有する別のネットワークエンティティ（例えば、ルータ、スイッチ、別のコンピュータ）との間の通信を配信する。言い換えると、本実施形態と互換性のあるイーサネット（R）ネットワークは、動作の全二重モードにおけるエンティティ間の通信を伝達する専用媒体として動作する。

【0021】

本実施形態は、あるネットワークエンティティから別のネットワークエンティティへと方向付けられたデータストリームを、複数の論理チャネルに分割またはストライピングすることによって、高速データ通信速度（例えば、10 G b p s ）を実現する。論理チャネルは、1つ以上の物理的リンクによって伝達され得る。例えば、単一の物理的リンクは、導電体または光導体上で論理チャネルを配信するために、周波数分割多重伝送方式（FDM）または波形分割多重伝送方式（WDM）を使用するように構成され得る。あるいは、2つ以上の別個の物理的導体が採用され得る。ある具体的な実施形態において、それぞれの論理チャネルは

、ファイバー束またはリボンの中の個々の光ファイバ線、または別個の無線信号など、別個の物理的導体によって運搬される。

【0022】

当業者にとって明白であるように、複数チャネルを介してデータストリームを分配またはストライピングすることで、データストリームは個々のチャネルの実質的合計で伝送され得る。

【0023】

図1は、IEEE規格802.3イーサネット(R)仕様と関連して、本発明の1実施形態がいかに見られ得るかを示す。参照符号130は、物理的層における既存のギガビットイーサネット(R)規格仕様(すなわち、規格802.3、1998年度版、仕様の3.5節に示される)を示す。既存のイーサネット(R)アーキテクチャにおいて、ギガビットPHY(物理的層デバイス)は、ギガビット媒体独立インターフェース(GMII)によってネットワークモデルのより高い層に連結される。

【0024】

図1はまた、参照符号110と120を付けた、本発明の実施形態を、アーキテクチャ130と即比較するに適した形態で示す。アーキテクチャ130と同様、これらの実施形態は、7層のISO/IEC参照モデルの物理的層で実施され得る。特に、「物理的分割」または「物理的セット合せ」のサブ層が、アーキテクチャ110のディストリビュータ/コレクタ100を含むよう定義され得る。

【0025】

以下の記載からより理解されるように、アーキテクチャ110は、複数チャネル上で、1つの通信を個々のチャネルの合計とほぼ同等の伝送速度で送信または受信するように構成される。一方、アーキテクチャ120は、アーキテクチャ110のほぼ総合的な速度で、1つの通信を1つのチャネル上で伝えるように構成される。

【0026】

以下に記載されるように、アーキテクチャ110のディストリビュータ/コレクタ100は、1つ以上の別個の要素を含み得る。特に、図1の実施形態におい

て、ディストリビュータ／コレクタ100は、通信の部分を複数の論理チャネルを介して広めるために、接続されたコンピュータシステムから送信された通信の分配機能を実施する。しかしながら、通信を受信するとき、ディストリビュータ／コレクタ100は、複数チャネルからデータを収集し、1つのデータストリームに再セットみ立てし、接続されたネットワークエンティティに（例えば、MACあるいは媒体アクセス制御層またはサブ層を介して）伝える。

【0027】

図1において、ディストリビュータ／コレクタ100は、10GMI102によってISO/IECモデルの調停サブ層およびより高い層／サブ層に連結され、2GMI104によって複数のPCS（物理的コーディングサブ層）に連結される。10GMI102と2GMI104は、以下に記載するように、いくつかの局面でアーキテクチャ120のGMIとは異なる。

【0028】

アーキテクチャ120における物理的層デバイスは、1秒あたり複数ギガビットの情報を伝送および受信するために、より高速で動作しなければならないことを除けば、アーキテクチャ130のPHY（すなわち、物理的コーディングサブ層、物理的媒体付属、物理的媒体依存型）に相当するエンティティを含むように見られ得る。アーキテクチャ110のPHYはまた、同様のエンティティ、およびディストリビュータ／コレクタ100を含み得る。図1において、アーキテクチャ110は4つの別個のPHYを含んでいるが、本発明の別の実施形態において任意の数が実施され得る。以下でより詳細に記載されるように、PHYの数は、本発明の1実施形態にしたがった、高速イーサネット（R）インターフェースデバイスによって採用される論理チャネルの数の決定における要因であり得る。

【0029】

アーキテクチャ130と同様、PHYの完全な詳細が、アーキテクチャ110および120に示され得ない。特に、（図1には描かれないが）アーキテクチャ130におけるPCSとPMA（物理的媒体付属）との間でエンコードされたデータを配信するTBI（10ビットインターフェース）はまた、以下に記載されるように、アーキテクチャ110と120において対応部分を有する。

【0030】

媒体106は、上述されたように、それぞれのPHYに連結された1つの物理的通信媒体からなってもよいし、またはそれぞれが異なるPHYに連結された、複数の別個の信号導体を含んでもよい。媒体106は、そのトポロジーがイーサネット（R）プロトコルと互換性を有し、以下に記載する本発明の種々の実施形態において規定される速度で信号を伝達することが可能なように、選択される。

【0031】

例示の実施形態において、10GMI102および各2GMI104の設計および動作は、IEEE802.3規格に記載されるGMIの全二重サブセットに基づく。図1に示される実施形態の動作中、ディストリビュータ／コレクタ100は、媒体アクセス制御（MAC）層から10GMI102を介して1Gbpsより速い速度で（例えば、例示の実施形態においては約10Gbpsまで）、フレームまたはパケットを受信する。同様に、ディストリビュータ／コレクタ100は、逆の方向に動作して、同じ転送速度で再構築されたフレームをMAC層に提供する。この伝送速度は、ディストリビュータ／コレクタ100をそれぞれのPCSに接続する2GMIインターフェースを介して、データが転送される速度の合計とほぼ同等である。それ故に、図1において、各2GMIは約2.5Gbpsの速度で動作し得る。

【0032】

「フレーム」および「パケット」という用語は、本明細書において交換可能に使用され得、概して、物理的層デバイス内のMAC層から受信され、またはMAC層に送信される情報単位を示す。「ミニフレーム」または「ミニパケット」という用語は、複数のチャネルのうち1つを介して送信されるフレームの断片または一部を記載するために使用され得る。

【0033】

図2は、本発明の1実施形態において、高速なイーサネット（R）インターフェースが、複数の論理チャネルを介してデータをストライピングすることを可能にするために、適したアーキテクチャのブロック図である。例示されたアーキテクチャは、複数の集積回路を介して、1つの集積回路またはASIC（特定用途

集積回路) 内で、あるいは1つ以上のプリント基板または他の同様のコンポーネント内で、完全に実施され得る。さらに、図2に関連して記載されるアーキテクチャは、媒体独立型であることが意図される。これは、複数の物理的層デバイスが、金属、光学、無線、またはその他にかかわらず、任意のタイプのイーサネット(R) ネットワークに接続し得ることを意味する。

【0034】

図2において、MAC (媒体アクセス制御) モジュール200は、高速イーサネット(R) インターフェースがインストールされるホストまたはクライアントコンピュータシステムの、物理的層とより高いネットワークプロトコル層との間の媒介として機能する。特に、MACモジュール200は、イーサネット(R) パケットを送受信し、より高いプロトコル層で動作するプロセスの代わりに、イーサネット(R) プロトコルを実行する。ネットワークインターフェースの分野の当業者は、MACモジュール200の設計、機能、動作に精通する。本発明の本実施形態におけるMACモジュール200は、イーサネット(R) ネットワークの既存のMACサブ層と同様に動作し、本発明の1実施形態を実施するMACサブ層および/またはより高い層ならびにサブ層に必要な任意の変更は、以下の記載から当業者にとって明白である。

【0035】

MACモジュール200は、10GMI I 202を介して、ディストリビュータ204およびコレクタ206に連結される。例示される実施形態において、10GMI I 202は、約10Gbpsのデータ速度で動作するように構成される。しかしながら、本発明の別の実施形態において、MACモジュール200とディストリビュータ204との間のインターフェース、およびMACモジュール200とコレクタ206との間のインターフェースは、他の速度で動作するように構成され得る。特に、本発明の1実施形態において、このインターフェースを介する、実質的に10Gbpsより遅い速度の(例えば、1Gbps、100Mbps、10Mbps、1Mbps) 情報の伝送を支えることによって、より低速なイーサネット(R) 構成を支える。そのように低速で動作するとき、本発明の1実施形態は全二重動作に限定され得ない。本発明の実施形態は、10GMI I

202および／または以下に記載されるその他のインターフェースを介する、データ転送の速度を上げることによって増強され得る。

【0036】

図2の実施形態において、10GMI I 202は、それぞれの方向に32本のデータラインを含み、MACモジュール200に、またはMACモジュール200から一度に4バイトを運搬し得る。それ故、10Gbpsを配信するためには、312.52MBdの通信速度が必要である。両端が使用される、156.26MHzで動作中のクロック信号は、必要なデータ転送速度を可能にする。同じクロック基準信号が、以下に記載される1つ以上のその他のインターフェースについて使用され得るか、または複数のクロックが採用され得る。

【0037】

ディストリビュータ204は、ホストコンピュータシステムから、媒体290に連結された別のエンティティに向けられたイーサネット(R)フレーム(例えば、パケット)上で動作する。逆方向に伝わるデータトラフィックについては、コレクタ206は、ユーザまたはホストコンピュータシステム上で動作するアプリケーション(例えば、プログラム、プロセス)のネットワークエンティティから受信したイーサネット(R)フレームを受信および再セットみ立てする。

【0038】

特に、ディストリビュータ204は、ホストコンピュータシステムとネットワークエンティティ間で確立された複数の論理チャネルを介して、MACモジュール200から受信した各フレームを分割または割り当てる。受信側エンティティ上のコレクタと共に動作する、ディストリビュータ204は、イーサネット(R)フレームまたはパケットが、個々のチャネルのいずれよりも速い速度で、フレームをエンティティに伝達するために、複数のイーサネット(R)チャネルを介してストライピングされることを可能とする。

【0039】

媒体290から受信されたトラフィックについて、コレクタ206は、複数のチャネルを介してストライピングされた各フレームを再構築する。本実施形態において、フレームストライピングがデータリンクレベルより下で起こるため、M

ACモジュール200は、現在構成されているものより高速でフレーム要素（例えば、バイト）を送受信する能力以外には、動作においてはほとんど変更を必要とし得ない。しかしながら、本発明の別の実施形態において、MACモジュール200および／または適用可能なネットワークプロトコルスタックにおいてより高いその他の層またはサブ層のさらなる改変が必要であり得る。

【0040】

フレーム要素が複数チャンネル間で分散または割り付けられる様態、およびフレームが再構築される様態は以下のセクションにおいて詳細に記載される。しかしながら、要約すると、個々のフレーム要素（例えば、バイト）は、複数（例えば、図2において示される実施形態においては4つ）の論理チャンネル間でラウンドロビンに基づいて分配される。それ故、それぞれのチャンネルは、1つの「ミニフレーム」または「ミニパケット」を運搬し、そのコンテンツは、受信側エンティティにてその他のミニフレームのコンテンツと再結合される。

【0041】

別のタイプのインターフェースがまた、図2に示され、そのうちの第1のタイプのインターフェースが2GMI I 208aとして示される。本発明の特定の実施形態において、このインターフェースの構成は、コンピュータシステムにとって利用可能な論理チャンネルの数によって決定され得るか、またはその数を決定し得る。例えば、2GMI I インターフェースは、各方向に8本のデータラインを含み、ディストリビュータ204および／またはコレクタ206を、1つの物理的層デバイスまたは物理的コーディングサブ層（PCS）に連結する。結合された2GMI I が10GMI I 202と同じデータ量を配信するために、2GMI I 208aを含む各2GMI I が、10GMI I 202と同じ通信速度で動作してもよい。10GMI I 202（例えば、156.26MHz）によって使用される同じクロック周波数が、再度両端でサンプリングされ、必要な312.52MBd通信速度を達成するために使用され得る。したがって、この実施形態の動作中、各2GMI I は、10GMI I 202上で運搬される情報の約1/Nを運搬し得る。ここでNはチャンネル数である。4つの論理チャンネルが示される例示された実施形態において、2GMI I 208aおよびその他の2GMI I は、それ

ぞれ、各方向に約2.5 Gbpsを配信する。

【0042】

本発明の1実施形態において、10 GMI I 202の最適なデータ転送速度を可能にするためには、各2 GMI Iが最高の効率またはほぼ最高の効率で（例えば、約2.5 Gbpsで）動作することが必要である。したがって、2 GMI I 208aまたは別の2 GMI Iが、データ運搬を止めるかまたは低下した方式で動作する場合には、本実施形態を採用するイーサネット（R）インターフェースは、動作を止め、エラー回復プロシージャに入るか、またはその他の診断措置または修正措置を取り得る。しかしながら、本発明の別の実施形態において、ディストリビュータ204およびコレクタ206は、（例えば、1つ以上の論理チャネル上でのデータ交換を停止することによって）より少ない論理チャネルを使用するように、動作を変更し得る。そうでない場合には、（例えば、1つ以上の論理チャネル上でのデータ交換を遅くすることによって）動作速度を減速させる。

【0043】

複数PCSモジュール（参照符号210a～210dによって示される）は、既存のギガビットイーサネット（R）インプリメンテーションと実質的に同じ状態でイーサネット（R）フレーム要素のコーディングを実施する。図2に示されるように、1つのPCSモジュールは、ディストリビュータ204およびコレクタ206に接続されたそれぞれの論理チャネルにセットみ込まれる。本発明の例示された実施形態において、PCSモジュールは、現在のIEEE 802.3ギガビットイーサネット（R）規格と同様に8B/10Bコーディングを実施する。それ故、ディストリビュータ204から受信されたそれぞれのバイトは、PCSモジュールによって、ネットワーク290を介して続いて合図される10ビットコードに変換される。受信側エンティティにおいて、PCSモジュールはチャネル上で受信したミニフレームをデコードし、取り戻されたバイトをコレクタに提供する。

【0044】

PCSモジュール210a～210dは、シリアライザ／デシリアライザ（Serializer／Deserializers）（SERDES）に連結さ

れる。ここでSERDESは、物理的媒体付属（PMA）デバイスと考えられ得、参照符号214a～214dによって示される。PCSモジュール210a～210dは、既存のギガビットイーサネット（R）アーキテクチャから調節され得る10ビットインターフェースによって、SERDESに連結される。しかしながら、例えば、新たな10ビットインターフェース（その中の1つが図2の2TBI212aとして示される）が、10GMII202および2GMII208aと同じ通信速度およびクロック速度を有し、既存のギガビットイーサネット（R）アーキテクチャのTBIの約2.5倍の速度で動作するよう構成される。別の実施形態において、イーサネット（R）インターフェースは本明細書に記載される伝送速度ぐらいの伝送速度で動作し、これにしたがい10GMII202、2GMII208aおよび2TBI212aの通信速度も変更され得る。図2の実施形態において、それぞれのSERDESは、おそらくPMD（物理的媒体依存型）モジュールを介して、媒体依存型インターフェース（MDI）によって、適したイーサネット（R）通信媒体に連結される。

【0045】

前に記載したように、本発明の1実施形態は、複数の論理チャネルを介してデータをストライピングすることによって高速のデータ転送速度（例えば、約10Gbps）を達成する。しかしながら、本発明の実施形態はまた、個々のチャネルを介して通信する高速のイーサネット（R）インターフェースと互換性を有する。しかしながら、必然的に、そのような個々のチャネルは、協同的に動作する複数のチャネルより高速のデータ転送で動作しなければならない。

【0046】

したがって、図2はまた、例示された実施形態がPCS250と協同するように拡張されて、複数の論理チャネルではなく単一のチャネルを介して媒体292と通信し得ることを示す。特に、PCS250は、10GMIIを介してMACモジュール200に連結され、10Gbpsを交換するために必要な速度で動作する適切なインターフェース上でSERDES254と通信する。SERDES254は動作の単一チャネルモードに必要な速度で動作するMDIを介して、媒体292に連結される。

【0047】

当業者であれば理解するように、複数のチャンネルにわたるデータのストライピングが、ネットワークプロトコルスタックの異なるレベルにて実施され得る。例えば、（例えば、802.3リンク集合と同様）MAC層の上で実施される場合、複数ネットワーク「フロー」または「会話」が分配且つ収集される必要があり、現在のイーサネット（R）インプリメンテーションに使用されるネットワークインターフェースハードウェアのほとんどすべてが複製される必要がある。さらに、そのような「フローストライピング」中の個々のフローの速度は、個々のチャンネルの速度に制限される。

【0048】

対照的に、本明細書において記載される本発明の1つ以上の実施形態は、ネットワークプロトコルスタックのより低いレベルでのネットワークデータのストライピングを実施する。特に、図2の実施形態において、ネットワークデータが複数の論理チャンネルを介して分岐する、（その後宛て先で再セットみ立てされる）ポイントは、MAC層の下（例えば、物理的層内）に位置する。これらの実施形態において、ストライピングは、個々のMACフレームまたはパケットのコンテンツでなされるため、物理的層のリソースのみが複製される必要がある。

【0049】

複数チャンネルを介してデータストリームをストライピングすることの利点の1つは、受信側エンティティのバッファ必要条件が減少することである。特に、それぞれのチャンネルは、データストリームの断片のみを受信し、そのチャンネルと他のチャンネルとの同期を取るに必要な程度だけ、バッファされればよい。また別の利点は、本発明の1実施形態によって達成された高速化した伝送速度が、それぞれの個々のチャンネルにおいて採用される少しずつの向上によって可能になることである。すなわち、1 Gbpsの代わりに10 Gbpsで実行するように、すべてのインターフェース要素の動作能力を増大させるのではなく、ほとんどの要素は10 Gbpsの分数でデータを処理することが可能であることしか必要としない。

【0050】

以下に記載の本発明の1つ以上の実施形態は、4つの論理チャネルを利用して、専用イーサネット（R）媒体を介して通信する。当業者は、これらの実施形態が、より多いチャネルまたはより少ないチャネルを使用するためにはいかに変更され得るかを容易に認識する。任意の複数のチャネル、すなわち2つ以上のチャネルの使用は、本発明の別の実施形態において考慮に入れられる。しかしながら、例えば、4つのチャネルの場合、それぞれのチャネルは約3.125 G B dの通信速度で動作し得、全体のデータ転送速度が10 G b p sに達することを可能にする。

【0051】

本発明の本実施形態において、複数のチャネル間の最大のスキュー（例えば、伝搬の遅れ）が特定される必要がある。スキューは相対的に大きくてもよいし小さくてもよいが、何らかの最大値が特定されなければならない。最大のスキューの予測を特定することによって、本実施形態は、以下に記載されるように、動作中に生じる実際のスキューが特定されたスキューより大きくならない限りは、適切に動作するように構成され得る。当業者であれば、適切な最大のスキューが、複数の論理チャネルにわたって生じる伝搬の遅れの差異、および／または論理チャネルが運搬されるリンクの異なる物理的または動作特徴を確認することによって決定され得ることを認識する。

【0052】

特定された最大のスキュー値で動作することの1つの利点は、データを第2ネットワークエンティティに送信する第1ネットワークエンティティのディストリビュータが、受信側エンティティで生じるスキューを考慮する必要がないことである（すなわち、受信側エンティティは「開ループ」として動作し得る）。受信側エンティティにおいて、バッファは1つ以上のチャネルに適用され得、実際のスキューを相殺する。バッファ量は、特定された最大のスキューに比例し得る。当業者であれば理解するように、最大のスキューの予測がネットワークセグメントの所望の長さから得られるか、または測定され得る。あるいは、特定の所望の最大のスキュー値は、ネットワークセグメントの最大の長さを決定し得る。

【0053】

本発明の1実施形態において、ディストリビュータ（例えば、図2のディストリビュータ204）は、MACモジュールまたは層からのバイトのストリーム（例えば、フレーム）を受け入れ、ラウンドロビン方式で個々のバイトをサブストリーム（例えば、ミニフレーム）に分配する。図2の実施形態に示されるように、4つのチャンネルは、4バイト幅（wide）の10GMI Iで実施され得る。したがって、ディストリビュータが4バイトを受信する毎に、1バイトが各チャンネルに提出される。この様態では、イーサネット（R）フレームは、異なるチャンネルにわたる伝送のために、4つのミニフレームに分割される。

【0054】

フレームの伝送は、4つのチャンネルのいずれで開始してもよいが、その後はフレームのバイトがラウンドロビン方式で分配される。すなわち、フレームの第1バイトは、チャンネルXに送信され得るが、その後チャンネルXはさらに、5、9、13バイトなどを運搬し得、その次の順序のチャンネルは、2、6、10バイトなどを運搬し得る。ちょうどフレームが任意のチャンネルで開始し得るのと同様に、フレームが終了するチャンネルは、フレームの長さによって決定される。本実施形態において、イーサネット（R）のフレーミング特徴が維持され、必要に応じて以下に記載されるように補足され得る。

【0055】

受信側エンティティにおいて、コレクタは連続的にそれぞれのチャンネルをモニタリングし、パケット間のアイドル期間中に受信した順序付け情報を使用して、チャンネルの同期を取ることを試みる。すべてのチャンネルの同期が取られ、コレクタがすべてのチャンネル上の同じフレームからミニフレームを受信するようになるまで、コレクタはアイドル状態をMACモジュールまたは層に報告する。いったんチャンネルの同期が取られ、同じパケットに属するデータを送達し始めると、コレクタは、またラウンドロビン方式で、それぞれのチャンネルから1度に1バイトを受け取り、バイトを再セットみ立てし、バイトストリームをMACに転送する。以下に記載されるように、それぞれのフレームおよびミニフレームの最初および最後のバイトは、容易に認識されるように印付けされる。

【0056】

すでに記載されたように、バッファは各チャネルのために採用され得、そしてチャネル間で予測される最悪のスキューに比例したサイズであり得る。それ故、実際のスキューが、バイト、いくつかのバイトの伝送または伝搬時間、あるいはミニフレーム全体の伝送または伝搬時間までも超える場合であっても、コレクタは依然首尾よくパケットを再セットみ立てし得る。

【0057】

フレームが複数のチャネルを介して分配される様態（例えば、1バイトずつ）のため、それぞれのミニフレームは、本実施形態において元のフレームの約4分の1からなる。これによって、受信側エンティティにおけるエラー検出の一意的な方法が可能となる。特に、カウンタが、特定のフレームについてそれぞれのチャネル上で受信されたバイト数を数えるために使用され得る。あるチャネル上で受信されたバイト数が、他のチャネル上で受信されたバイト数より1より多く異なる場合、エラーが起こったと判定され得る。例えば、MACに無効なフレームの受信を通知することによって、エラー訂正が次いで開始され得る。

【0058】

ディストリビュータからPCSが受信したそれぞれのミニフレームは、完全なギガビットイーサネット（R）パケットが従来のギガビットイーサネット（R）インプリメンテーションにおいてフレームそしてエンコードされた方法と同様の様態で、「フレーム」およびエンコードされる。特に、本発明の1実施形態において、PCSモジュールは、8B/10Bコーディング方式を適用して、ディストリビュータまたは物理的リンクのそれぞれから受信されたそれぞれのデータサブストリームをエンコードまたはデコードする。その他のコーディング方式（例えば、4B/5B、NRZIなど）は、本発明の別の実施形態において使用され得る。しかしながら、本実施形態のアーキテクチャのために、いくつかの改変がコーディング方式にとって必要であるかもしれない。

【0059】

例えば、4つのチャネル間でフレームバイトのラウンドロビン分配を行うことによってあるチャネルは、フレームのプリアンブルフィールド（通常長さが7バイトである）から1バイトしか受信しないことになる。特に、既存のイーサネッ

ト(R)アーキテクチャにおいて、それぞれのフレームのプリアンブルフィールドの1バイトは、エンコーディング中にパケットデリミッターの開始(SPD)の記号によって置き換えられる。さらに、パケット間ギャップ(IPG)は、ギャップのそれぞれのアイドル記号が1セットの2つのコードに変換されるようにエンコードされる。したがって、新たなフレームまたはミニフレームのタイミングに依存して、チャンネルのミニフレームがおそらく最初のプリアンブルバイトを失う。これはアイドル(すなわち、第2アイドルコード)の送信を終了する必要ゆえである。チャンネルが1プリアンブルバイトしか有さず、アイドル延長にそのプリアンブルバイトを奪われた場合、チャンネルはSPD記号によって置き換えられ得るプリアンブルバイトを有さない。この問題に対する1つの解決法は、アイドル記号がそのプリアンブルを犠牲にする複数のコードを必要としないように、コーディング方式を変更することである。別の解決法は、MACによって生成されるプリアンブルのサイズを8(またはそれより多い)バイトにまで大きくすることである。しかしながらその他の解決法が当業者によって認識され得る。

【0060】

本発明の1実施形態(例えば、図2の実施形態)を実施することによるまた別の影響は、複数チャンネル間でIPG(通常最小で12バイトである)を分配するときに生じる。図2の実施形態において、例えば、最小サイズのIPGによって、3バイトの各チャンネルの、ミニフレーム間ギャップを生じる。現在のコーディング方式によって、3バイト/コードまでのパケットデリミッターの終了(EPD)が可能になる。それ故、最大サイズEPDが最小サイズIPGと共に使用される場合、チャンネルは同期を取るべきいかなるアイドルコードも受信し得ない。この状況に対する解決法の中には、1つのコードのみ、またはせいぜい2つのコードからなるEPDの使用がある。別の解決法は、最小IPGのサイズを大きくすることである。

【0061】

コレクタによるチャンネルの同期化を助けるために、本発明の1実施形態において、いくつかの列挙されたアイドル記号が適用される。これらの記号はアイドル1、アイドル2、・・・、アイドルNとして表され得る。異なるアイドル記号の

数はコーディング方式によって制限され得るが、64の範囲または128の範囲ですら、本発明の別の実施形態において考慮に入れられる。例えば、同じアイドル記号が複数チャネルのそれぞれを介して伝送され、そしてそれぞれのMACフレームとともに変化する。それ故、第1MACフレームと第2MACフレームの間のIPGはアイドルXで印付けされ得、第2と第3フレームの間のIPGはアイドルX+1などで印付けされ得る。

【0062】

さらに、各フレームおよびミニフレームの初めと終わりを効果的に区切るために、デリミッターのさらなるセットが本発明の1実施形態において適用される。本実施形態において、パケットデリミッターの開始（SPD）およびパケットデリミッターの終了（EPD）は、ディストリビュータにてMAC層から受信された各パケットの初めと終わりにそれぞれ挿入される。それ故、SPDおよびEPD記号は、既存のイーサネット（R）アーキテクチャと同様に使用され得る。ミニフレームデリミッターの開始（SMD）およびミニフレームデリミッターの終了（EMD）と呼ばれ得る新たなデリミッターのセットが、SPDまたはEPD記号で印付けされないミニフレームのそれぞれの初めと終わりを印付けするために使用される。それ故、いずれのチャネルでパケットが開始または終了するかにかかわらず、パケットを開始するミニフレームは、SPDコードで開始し、パケットを完了するミニフレームはEPDコードで終了する。その他のミニフレームは、SMDコードで開始し、EMDコードで終了する。

【0063】

図3A～3Bは、本発明の1実施形態において、複数のチャネルを介してパケットを伝送する1方法およびパケットを受信する1方法を示すフローチャートである。図3A～3Bが示す方法を実施するために、上述のようなイーサネット（R）インターフェースデバイスは、4つの論理チャネルを介して各パケットをストライピングすることによって、約10Gbpsの速度でそれぞれの方向にデータを伝送および受信するように構成される。

【0064】

状態300は、図3Aの開始状態である。状態302は、アイドル状態であり

、MAC層またはモジュールからイーサネット（R）インターフェースデバイスのディストリビュータに流れるパケットデータの欠如によって特徴付けられる。ディストリビュータは、4つのチャネルのそれぞれに、適切なアイドル記号またはバイトを伝送することによって、アイドル状態を示す。しかしながら、特に、ディストリビュータは、同じアイドル記号をそれぞれのチャネルのPCSに送信し、ここでアイドル記号はいくつかの異なる記号の中の1つである。概して、同じアイドル記号が、同時にそれぞれのチャネルで送信されるが、いずれの記号がそれぞれのアイドル時間中に送信されるかを変えることによって、受信側のイーサネット（R）インターフェースデバイスのコレクタは、より容易にチャネルの同期を取ることができる。PCSによってそれぞれのアイドル記号が受信されると、PCSは10ビットコードとして記号をエンコードし、適切なイーサネット（R）媒体を介して合図するためにその記号を転送する。

【0065】

状態304において、ディストリビュータはMACからのパケットの受信を開始する。ディストリビュータは、TX__EN信号ラインの状態の変化によって、パケットの開始を検出し得る。この実施形態において、MACとディストリビュータとを連結するインターフェースは、32データビット幅であって、したがって、約10Gbpsの速度で1度に4バイトまでを送達する。それ故、この実施形態において、4バイトのセットがMACから受信される毎に、1バイトがチャネルを介して伝送され得る。

【0066】

状態306において、ディストリビュータは、エンコーディングのためにそれぞれのチャネルのミニフレームの第1バイトをPCSに送信する。

【0067】

状態308において、それぞれのPCSは、特定のコードで第1バイトをエンコードする。特に、全体のパケットの第1バイトは、いずれのチャネルまたはミニフレームにまたがっているにもかかわらず、受信局によって理解されるコードに変換されて、新たなパケットの開始を示す。その他のチャネルの開始バイトは、（異なるコードで）同様にエンコードされて、新たなミニフレームの開始バイ

トとしてそのステータスを示す。

【0068】

状態310において、パケットの残りは、ディストリビュータによって受信され、それぞれのチャンネルに（ラウンドロビン方式で）1度に1バイトずつ分配され、エンコードされ、伝送される。

【0069】

状態312において、パケットの最後のバイトを含む、4つのミニフレームの最後のバイトはまた、エンティティを受信することによって認識される特定コードに変換される。特に、パケットの最後のバイトを運搬しない各ミニフレームの最後のバイトは、第1終了コードでエンコードされ、全体のパケットの最後のバイトがその他の特有のコードでエンコードされる。次いで例示されたプロシージャは、状態314で終了する。

【0070】

図3Bにおいて、図3Aのプロシージャで送信されたパケットを受信する1プロシージャが示される。図3Bにおいて、状態350は開始状態である。状態352は、アイドル状態であり、これは、図3Aに示されたパケットを伝送するために使用されたイーサネット（R）媒体と同じイーサネット（R）媒体に連結された受信側エンティティのコレクタが、媒体を介してパケットを受信しないことを意味する。特に、コレクタが通信チャンネルの同期を取ることができない限り（例えば、4つのチャンネルのそれぞれで少なくとも1つ同じアイドルコードを受信する）、コレクタはあたかもどんなトラフィックも受信していないかのように動作し得る。

【0071】

しかしながら、状態354において、コレクタは、4つのチャンネルすべてで同じアイドルコードを検出することによってチャンネルの同期を取り得る。先に記載したように、弾性バッファは1つ以上のチャンネルで採用され得、チャンネルのキューまたは他の伝搬の遅れを補償する。コレクタは同期を取り、次に、それぞれのチャンネルが1つのパケットの部分の送達を開始することを待つ。

【0072】

状態356において、受信デバイスにおけるそれぞれのチャンネルのための物理的コーディングサブ層は、伝送側エンティティから送信されるミニフレームの第1コードを受信する。それぞれのミニフレームの第1コードは、上に記載したように、それぞれのコードに特有のコードによって認識される。いずれのチャンネルでパケットの第1バイトが受信されるかを決定することによって、コレクタは、パケットの残りのバイトを、（ラウンドロビン方式で）読み出す適切な順序を決定し得る。

【0073】

それ故、状態358～360において、コレクタはそれぞれのチャンネルで1度に1バイトを受信し、それを適切な順序でMACに転送する。したがって、パケットのコンテンツは、送信側のMACによってディスパッチされた順序と同じ順序で受信側のMACに到達する。

【0074】

状態362において、それぞれのミニフレームの最後のバイトおよびパケットの最後のバイトは、それらの特有のコードによって認識される。例えば、ちょうど送信側イーサネット（R）インターフェースデバイスにあるPCSモジュールが、アイドル記号の代わりに終了デリミッターを構築するように、受信デバイスのPCSモジュールは、終了デリミッターをアイドル記号に戻すように変換し得る。説明されたプロシージャは次いで、状態364で終了する。

【0075】

図4は、本発明の1実施形態における、図2の10GMII202にわたる、64バイトの長さ（例えば、60データバイトプラス4CRC（周期的冗長検査）バイト）のパケットの伝送、その後の、65バイトの長さの複数のパケットの伝送を示す。図4で示されるその他の3つのバスは、クロックバス、Transmit_Enable（TX_EN）／Receive_Data_Valid（RX_DV）バス、そしてValid（VLD）バスである。当業者であれば理解するように、TX_ENバスは、MAC層と、パケットを伝送する第1ネットワークエンティティにあるディストリビュータとを連結し、RX_DVバスは、MAC層と、パケットの受信側にある第2ネットワークエンティティにあるコ

レクタとを連結する。衝突およびキャリア感知信号は、全二重モードで動作するため、例示された実施形態に含まれない。

【0076】

双方向に動作するVLDバスは、MACモジュール200からディストリビュータ204に、またはコレクタ206からMACモジュール200に、データバス（例えば、10GMI I 202）を介して転送される有効なバイト数を示す。VLDバスは、（データ転送の方向に依存する）TX__ENバスまたはRX__DVバスの状態と共にその状態を解釈することによって、二本のラインの幅に制限され得る。特に、データバスについての以下の記載からよりよく理解されるように、VLDバス上の非0値は、TX__ENまたはRX__DVがアサートされる場合有意義である。これらのバスのいずれかがアサートされる場合、VLDバス上の0値は、4つの有効なバイトがデータバスを移行していることを示す。そうでない場合には、VLD上の0値は、データバスがアイドルである（すなわち、データを運搬しない）ことを示す。

【0077】

10GMI I 202では、4バイトが1度に伝達される。それ故、図4における時間 t_1 において、第1パケットの最初の4つのプリアンブルバイトが送信され、時間 t_2 において、他の3つのプリアンブルフィールドバイトおよびフレームデリミッターの開始（SFD）記号が送信され、時間 t_3 において、最初の4つのデータバイトが送信され、などである。

【0078】

図4は、クロック信号の両端のデータの転送を示す。TX__EN/RX__DVおよびデータバスと共にVLDバスを検討することによって、MACフレームの初めおよび終わりで、VLDバスがいかに、0値から非0値に移行し、0値に戻り得るかが分かる。

【0079】

図5A～5Dは、図4に示されたフレームの、本発明の1実施形態による別個のチャンネルにわたる伝送のための複数のミニフレームへの変換を示す。特に、図4において10GMI I 202を介してMACからディストリビュータに伝達さ

れたデータストリームは、図5A～5Dにおいて2GMII208a、208b、208cおよび208dを介して分配される。さらに、2TBI212a、212b、212cおよび212dは、各PCSからエンコードされたバイトを配信する。参照のために、（図4における周波数と同じ周波数で動作する）クロック信号、TX__EN/RX__DVバスおよびTX__ER（伝送エラー）/RX__ER（受信エラー）バスがまた、図5A～5Dに示される。

【0080】

図5A～5Dに示されるように、それぞれの2GMIIは8ビットの幅であり、そのクロック信号の両端はデータ転送のために使用され、衝突およびキャリア感知信号は、この実施形態が全二重動作であるために省略され得る。パケットデリミッター（PD）信号は、MACフレームの最初と最後のバイトを特定するために、それぞれの方向に（すなわち、ディストリビュータから各PCSに、そして各PCSからコレクタに）追加される。それ故、パケットの開始は、PDおよびTX__EN信号を上げることによって合図され得、そしてパケットの終了は、同じ信号を下げることによって合図され得る。それぞれの2TBIは10ビット長であって、クロック信号の両端は再度データ転送のために使用される。

【0081】

例示のために、図5A～5Dにおいて、バイトを移行する2GMIIバスは、図4とはやや異なって示される。具体的には、IPGコードまたはアイドルは文字「I」によって表され、PA（プリアンプル）は文字「P」、CRCは「C」によって示される。これらの文字のそれぞれは、順に増える数によって変更される。それ故、フレームの7つのプリアンプルバイト、4つのCRCバイト、そして種々のアイドル記号が容易に示され得る。

【0082】

図5A～5Dにおける各ミニフレームは、同一のアイドル記号（例えば、第1パケットの前のアイドル1）に先行される。例えば、それぞれの後続のパケットが伝達された後、異なるアイドル記号がパケット間ギャップのために使用される。それ故、図5A～5Dの実施形態において、最低4つの異なるアイドル記号が必要なコーディング方式が採択される。

【0083】

本発明の種々の実施形態のエラー検出および処理能力は、上記アーキテクチャの独創的な特性を利用し得る。例えば、パケットを含むミニフレームは、1バイトより多くは長さが異ならないので、コレクタはミニフレームの長さを比較することによって無効なパケットを検出し得る。さらに、チャンネルのスキューが（例えば、特定された最大のスキューの予測によって）制限されるので、チャンネルバッファがオーバーフローする場合、チャンネルまたは物理的リンクが欠陥を有しているかまたは仕様外であるか、あるいは何らかの他のエラーが生じて、ミニフレームの遅れまたは崩壊がもたらされている可能性がある。

【0084】

チャンネルの同期化エラーは、パケット間で合図されたシーケンス情報（例えば、異なるアイドルコード）を使用して、コレクタによって検出され得る。採用される異なるアイドルコード数が多ければ多いほど、シーケンシングエラーが検出されずに伝えられるためにチャンネルで損失または投入されたであろう連続ミニフレームの数も多くなる。十分に多い種々のアイドルコードによって、チャンネルバッファは、同期化エラーがデータのフローに影響を及ぼし得る前にオーバーフローし得、それ故に別のレベルのエラー抵抗を提供する。

【0085】

パケットデータの崩壊をもたらす個々のビットエラーは、コレクタによるミニフレームの再セットみ立て後に、MACレベルにて（例えば、CRC演算によって）検出および処理される。コーディング違反、フレーミングエラー、不一致エラーなどに関連するその他のエラーは、PCSレベルにて検出され得る。特に、コレクタで受信されるパケット（例えばミニフレームのセット）毎に、コレクタは、（例えば、それぞれのPCSを介して）任意のパケットのミニフレームを処理する際にエラーが検出されたか否かを通知され得る。それ故、パケットの1つのミニフレームにおけるエラーはパケット全体にその原因を帰する。

【0086】

チャンネル内の複数のミニフレーズの損失または挿入など、コレクタによって検出されないチャンネルの同期化エラーは、MACによって検出される。これは、チ

チャネルの同期化エラーが、（例えば、フレーミング、コーディング、パリティなど）他のエラーのない非常に多い数のCRCエラーを生じ得るからである。これらのタイプのエラーの回復は、短期間を介してリモートエンドからの伝送を止める、リンク再初期化または802.3フロー制御の使用を含み得る。これによってすべてのチャネルが自動的に再度同期を取る。

【0087】

本発明の実施形態の上記記載は、例示および説明の目的のみにために提示される。本発明の上記記載は、本発明のすべての実施形態を網羅するものではなく、開示される形態に本発明を限定するものではない。多くの変更および改変が当業者にとって明白である。したがって、上記開示は、本発明を限定するものではない。本発明の範囲は、上掲の特許請求の範囲によって規定される。

【図面の簡単な説明】

【図1】

図1は、既存のギガビットインサネットアーキテクチャに関連して、本発明の1実施形態の機能の概念的な層の構造(layering)を示す図である。

【図2】

図2は、本発明の1実施形態を含むイーサネット(R)ネットワークインターフェースデバイスの一部のブロック図である。

【図3A】

図3Aは、本発明の1実施形態による、複数チャネルを介してパケットを分配する1様態を例示するフローチャートである。

【図3B】

図3Bは、本発明の1実施形態による、複数チャネルを介して伝送されるパケットを収集する1様態を例示するフローチャートである。

【図4】

図4は、本発明の1実施形態による、1秒あたり複数ギガビットのインターフェースでの、複数のイーサネット(R)フレームを含むデータストリームの転送を示す図である。

【図5A】

図5 Aは、本発明の1実施形態による、複数チャンネルにわたる、図4のデータストリームの分割を示す図である。

【図5 B】

図5 Bは、本発明の1実施形態による、複数チャンネルにわたる、図4のデータストリームの分割を示す図である。

【図5 C】

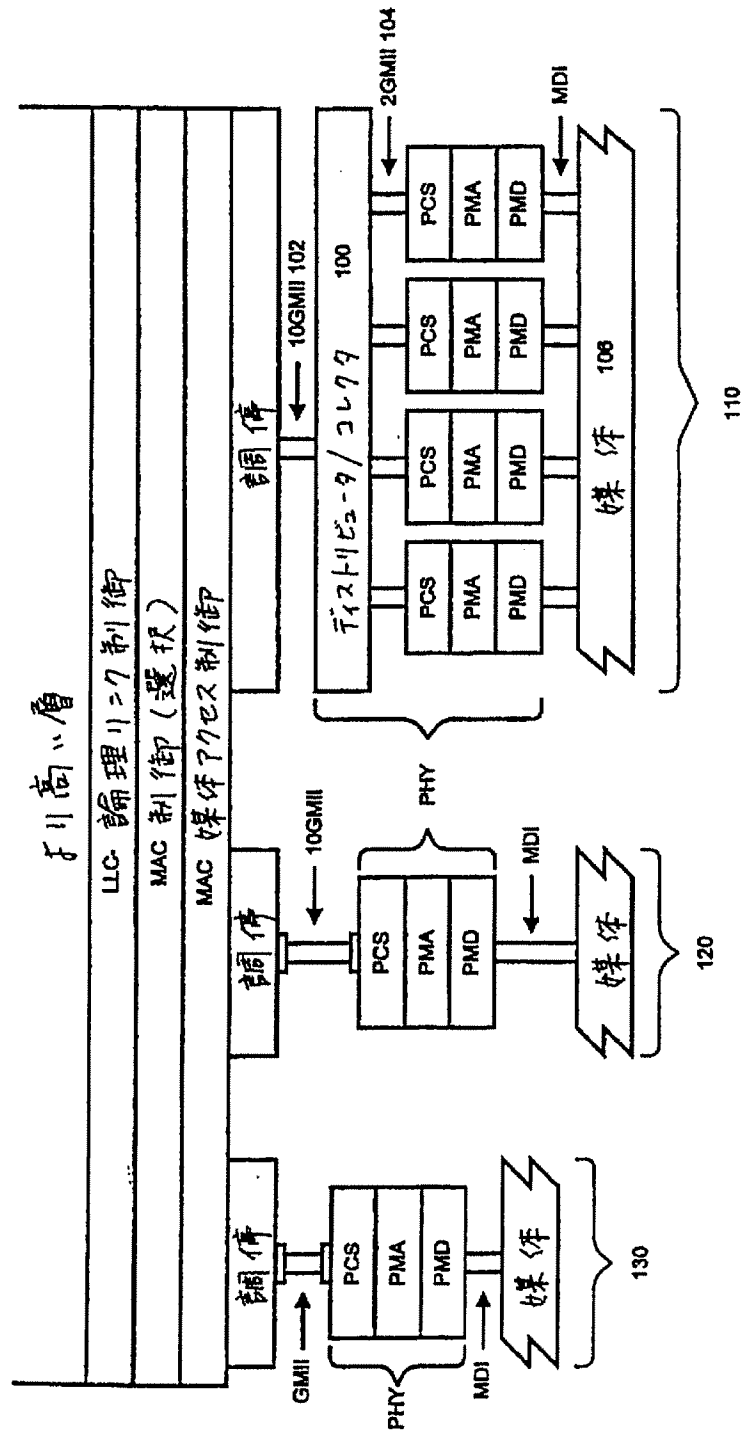
図5 Cは、本発明の1実施形態による、複数チャンネルにわたる、図4のデータストリームの分割を示す図である。

【図5 D】

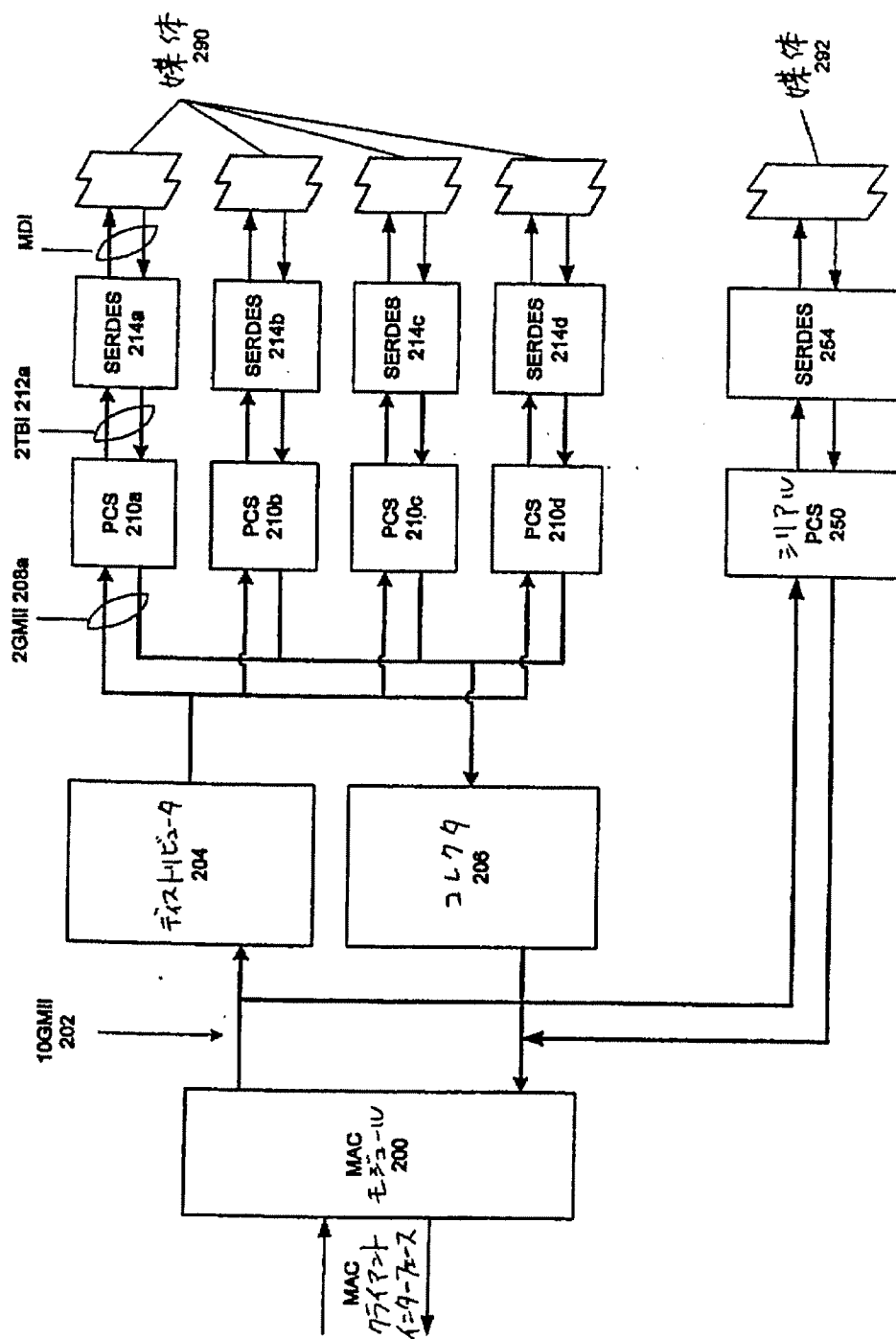
図5 Dは、本発明の1実施形態による、複数チャンネルにわたる、図4のデータストリームの分割を示す図である。

【符号の説明】

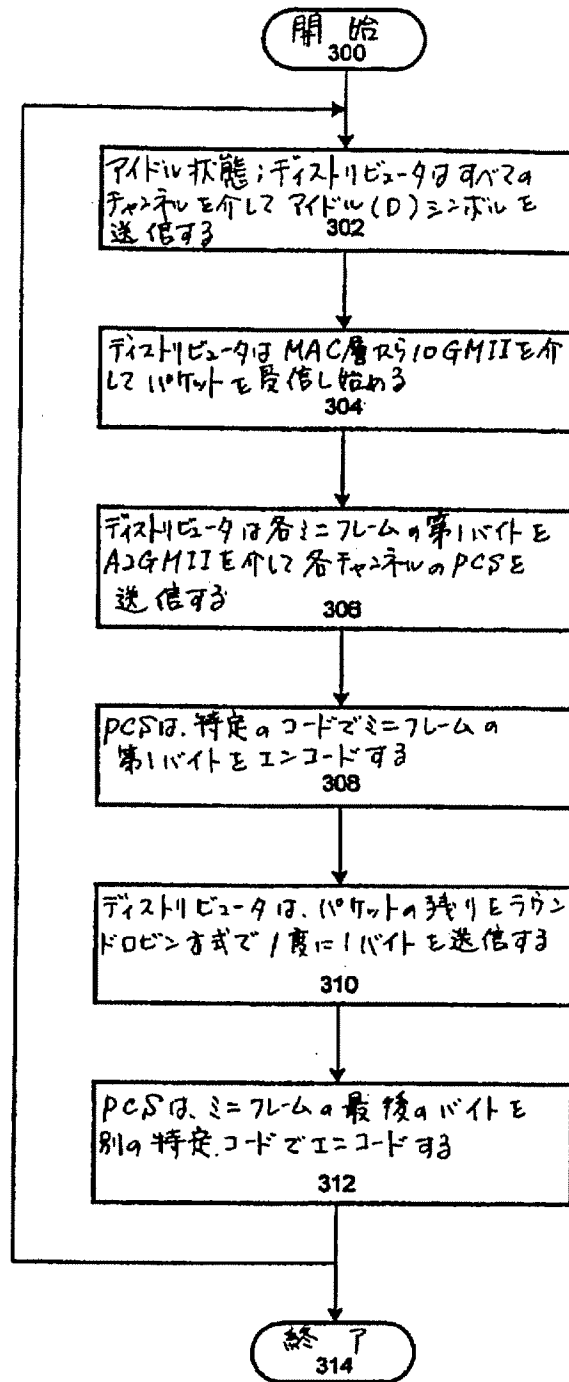
【図1】



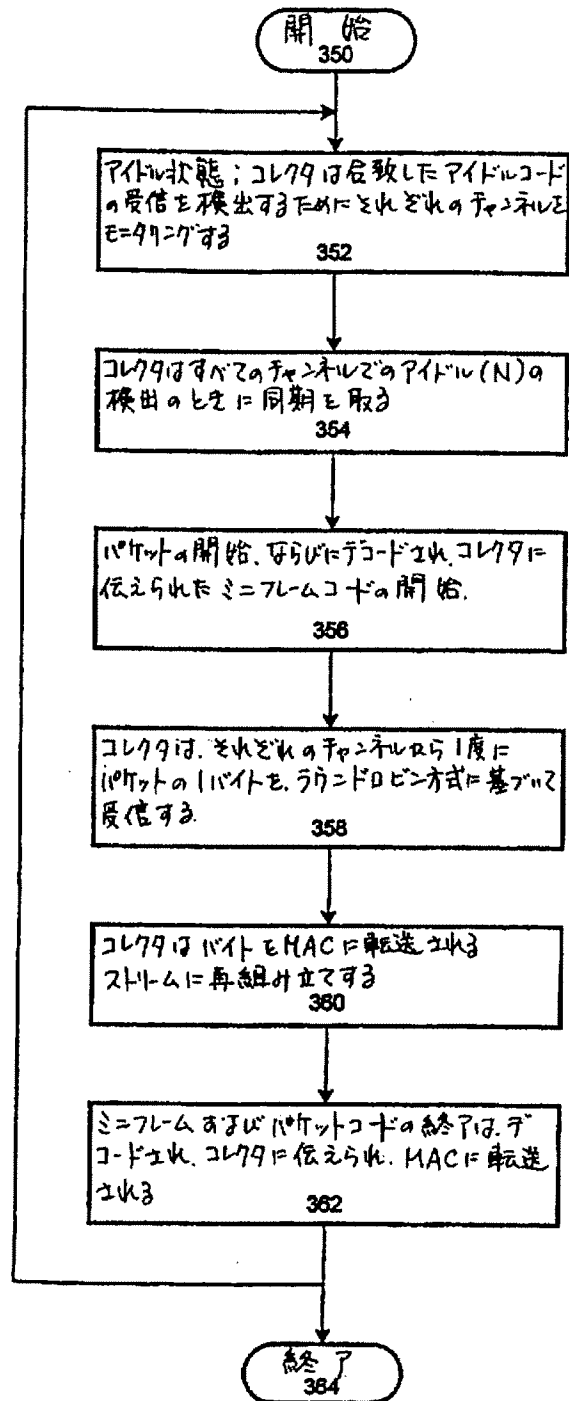
【図 2】



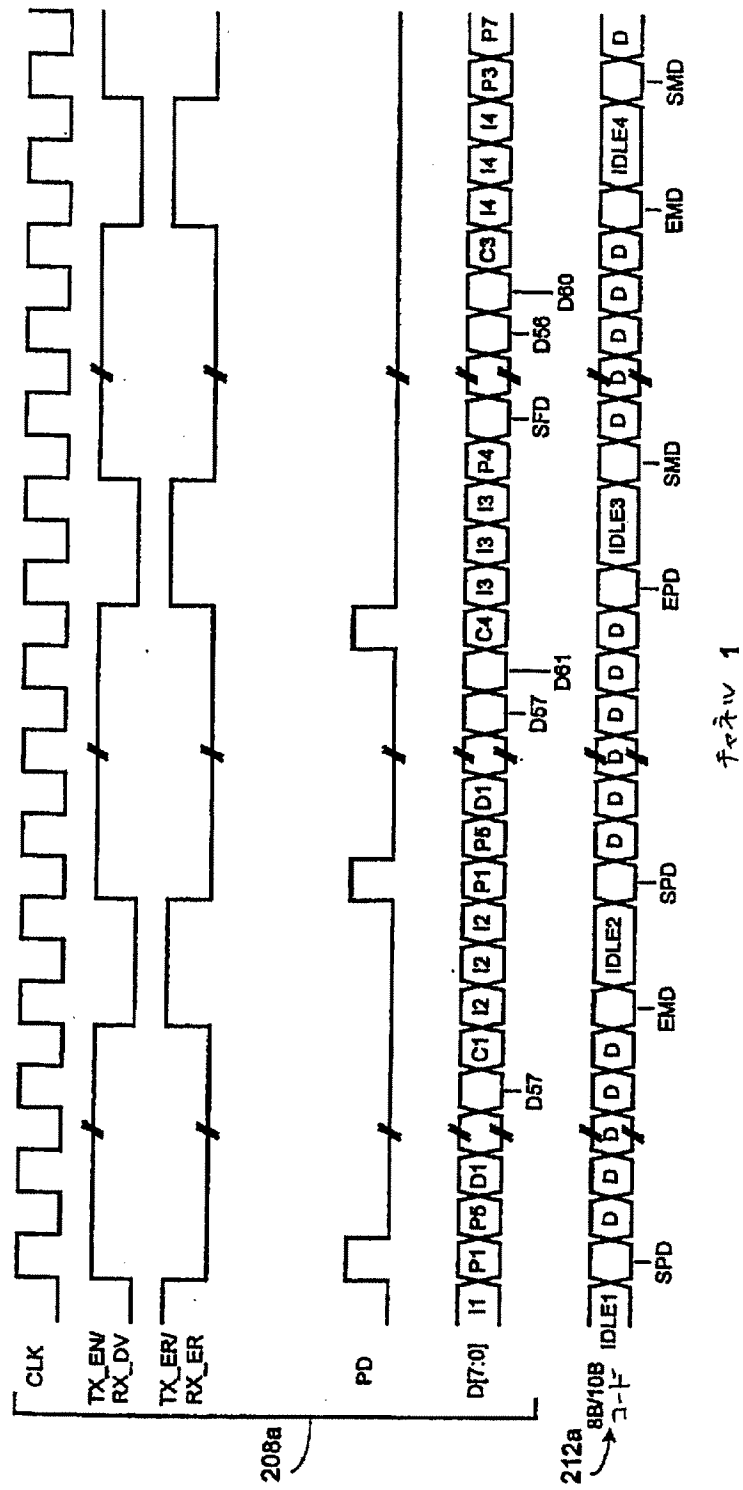
【図3A】



【図3B】



【図 5 A】



【図5B】

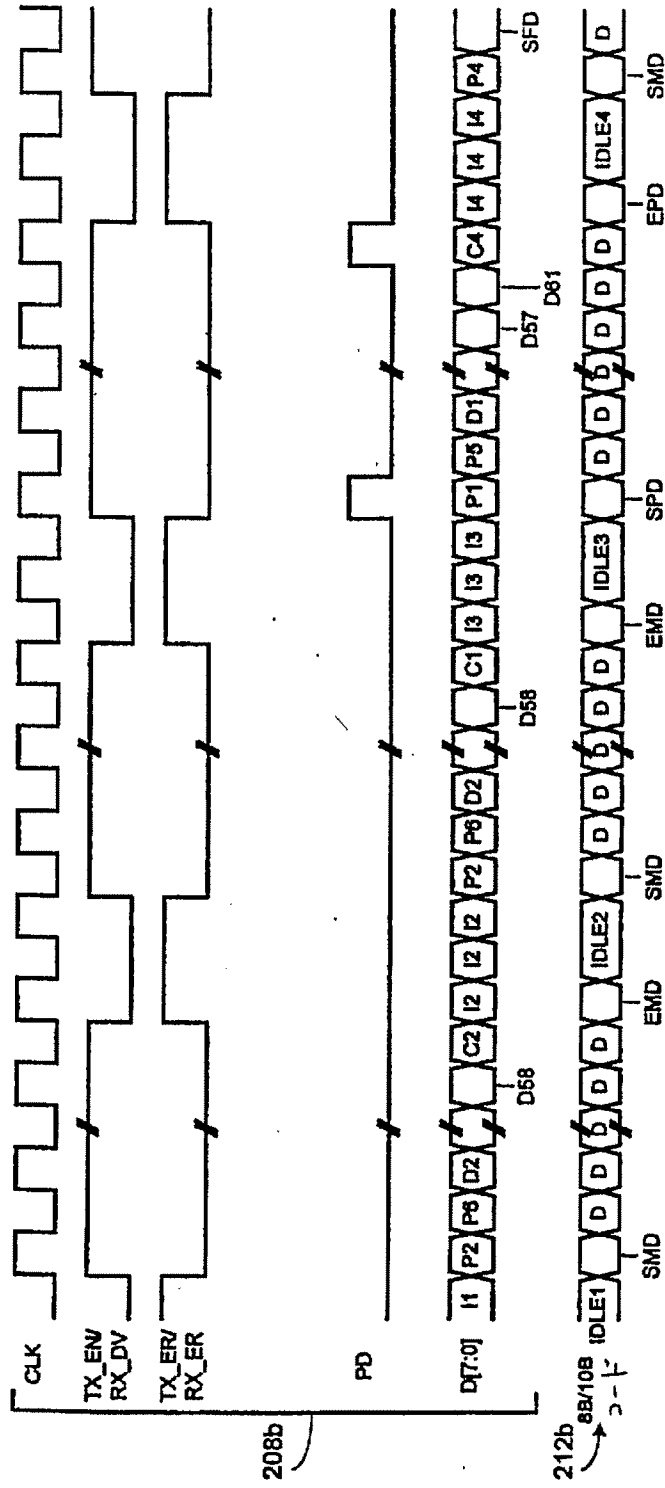


図5B 2

【図5D】

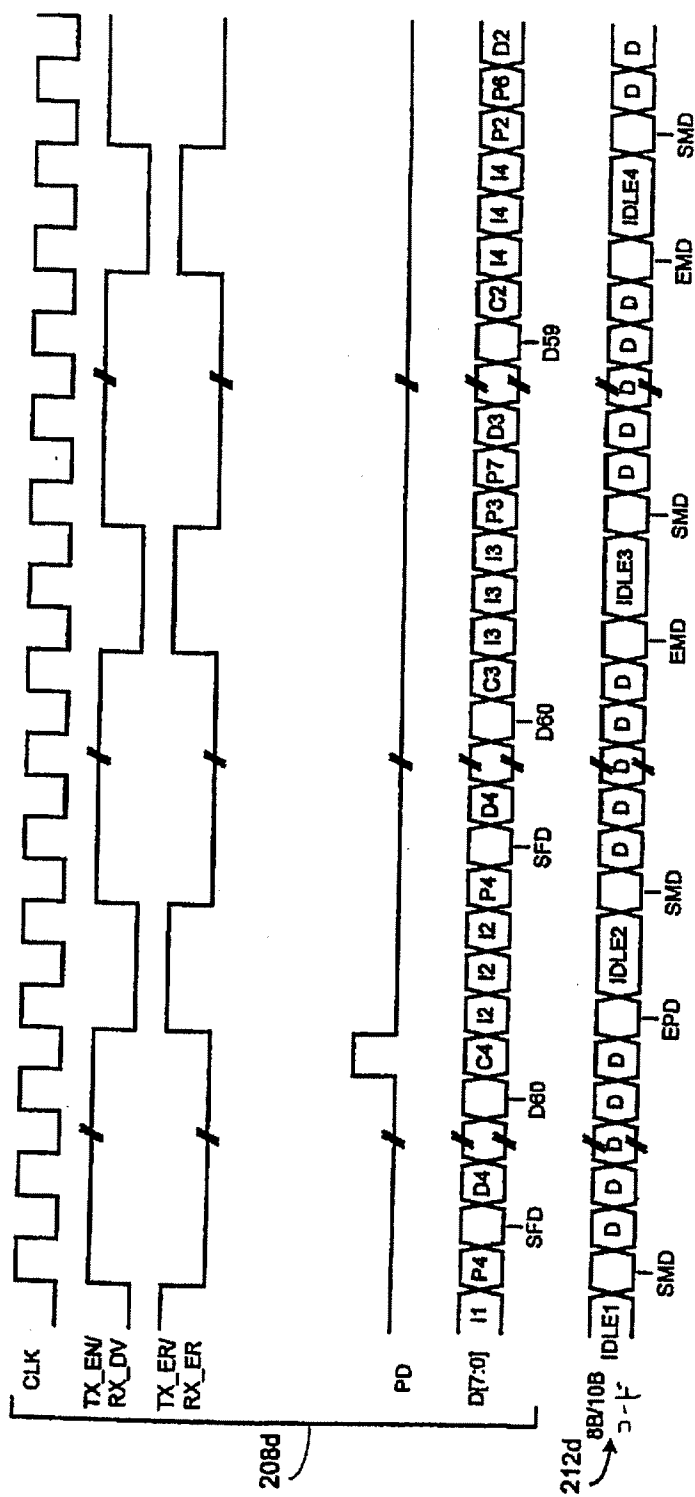


Figure 4

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 00/13584	
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 H04L12/413	
According to International Patent Classification (IPC) or to both national classification and IPC	
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 H04L	
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched	
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, PAJ, INSPEC	
C. DOCUMENTS CONSIDERED TO BE RELEVANT	
Category *	Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No.
P,X	GB 2 332 128 A (ADVANCED MICRO DEVICES INC) 9 June 1999 (1999-06-09)
P,A	page 5, line 5 -page 10, line 26 --- -/---
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C	
<input checked="" type="checkbox"/> Patent family members are listed in annex	
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document relating to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family	
Date of the actual completion of the international search 17 January 2001	Date of mailing of the international search report 26/01/2001
Name and mailing address of the ISA European Patent Office, P.O. 5618 Patentlaan 2 NL - 2200 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl Fac. (+31-70) 340-3016	Authorized officer RAMIREZ DE AREL., F

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 00/13584

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 438 571 A (CROUCH SIMON E ET AL) 1 August 1995 (1995-08-01)	1,3,4,6, 9,10,16, 17, 20-23, 26,28, 30,31, 33,34,38
A	column 4, line 9 -column 6, line 63 column 11, line 32 -column 14, line 19	2,5,7,8, 11-14, 18,24, 25,27, 29,32, 35,39
X	US 5 640 605 A (JOHNSON HOWARD W ET AL) 17 June 1997 (1997-06-17)	1,3,4,9, 10,16, 17, 20-23, 30,31,33
A	column 3, line 35 -column 5, line 41	2,5,6, 11-14, 18, 24-26, 28,32, 34,38,39
A	ZIMMERMAN C ET AL: "TRUNKING BRANCHES OUT" DATA COMMUNICATIONS,US,MCGRAW HILL. NEW YORK, vol. 27, no. 18, December 1998 (1998-12), pages 62-66,68-69, XPO00669883 ISSN: 0363-6399 the whole document	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 00/13584

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
GB 2332128 A	09-06-1999	US 6081523 A	27-06-2000
		JP 11187051 A	09-07-1999
US 5438571 A	01-08-1995	US 5550836 A	27-08-1996
		JP 7321839 A	08-12-1995
		US 5598406 A	28-01-1997
		CA 2101860 A	07-05-1994
		EP 0596523 A	11-05-1994
		EP 0714191 A	29-05-1996
		JP 6216925 A	05-08-1994
US 5640605 A	17-06-1997	US 5583872 A	10-12-1996
		AU 3366995 A	22-03-1996
		EP 0777876 A	11-06-1997
		WO 9607132 A	07-03-1996

フロントページの続き

(81)指定国 EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG), AP(GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW

(72)発明者 ヘンデル, アリエル

アメリカ合衆国 カリフォルニア 95014,
キューパーティーン, ニューキャッスル
ドライブ 7537

Fターム(参考) 5K032 AA02 AA09 CA06 CC10 CC11
CC13 DA11 DB18
5K033 AA02 AA09 CA06 CB14 CB15
DB11